File S5: Derivation of expressions to quantify the costs or benefits of misregulation.

In this section I derive multiple expressions from the main text that quantify fitness and the proportion of correctly expressed genes in a new state when misregulation in the old state is present compared to when it is absent. I first begin with the simple but illustrative case that the two random variables $g_i^N = 1$ and $g_i^O = 1$ are statistically independent. In this case, the joint probability $P(g_i^N = 1, g_i^O = 1)$ takes on the very simple form $P(g_i^N = 1, g_i^O = 1) = P(g_i^N = 1|g_i^O = 1)P(g_i^O = 1) = P(g_i^N = 1)f_1^O = f^N f_1^O$, where the symbol $'|'$ denotes a conditional probability. This identity helps compute the fraction f_{11}^N of correctly active genes in the new state. Specifically, if we apply the above argument to all genes g_i, then $f_{11}^N = f_1^O f^N$. To see this, consider that among all those genes Gf^N that should be expressed in the new state, a fraction f_1^O is already expressed in the old state. With the notation of equations (1a)-(1d) from the main text, we then get

$$f_{11}^N = f_1^O f^N = (f_{11}^O + f_{01}^O) f^N \tag{8}$$

Analogously, the fraction of genes that are correctly off computes as

$$f_{00}^N = (1 - f_1^O)(1 - f^N) = (f_{00}^O + f_{10}^O)(1 - f^N) \tag{9}$$

As for the genes misexpressed in the new state, the fraction of wrongly active genes is given by

$$f_{01}^N = f_1^O (1 - f^N) = (f_{11}^O + f_{01}^O)(1 - f^N), \tag{10}$$

and the fraction of wrongly inactive genes as

$$f_{10}^N = (1 - f_1^O) f^N = (f_{00}^O + f_{10}^O) f^N \tag{11}$$

These associations assume that the genes that are expressed under optimal adaptation in the new state are stochastically independent from those expressed under optimal adaptation in the old state. However, this is not generally the case. A gene that is expressed in the old state is more likely to be expressed in the new state than expected by chance alone. Pertinent evidence includes that many genes are housekeeping genes and are expressed in all or most cell states (Eisenberg and Levanon, 2013; Wang et al., 2019). Also, microarray experiments that quantify gene expression in different environments and physiological conditions to which organisms are well-adapted show that only a modest fraction of genes (<1-20 percent) typically change their expression in two different organismal states (Dragosits et al., 2013; Colbourne et al., 2011; Gasch et al., 2000; Henry et al., 2012; Huang et al., 2008; Landis et al., 2004). Furthermore, RNA sequencing studies that quantify gene expression in different tissues or cell types from the same organism (Cardoso-Moreira et al., 2019; Wang et al.,

2019; Uhlen et al., 2015) show that the identity of genes expressed in different tissues is positively associated. (File S7, Figure S1B and S1C show examples from humans and mouse.)

For gene expression states that are correlated, the above expressions need to be modified. Consider first the extreme case where gene expression in the two states is perfectly correlated, that is, any gene that is (not) expressed in the old state needs (not) be expressed in the new state. In mathematical terms, this means that $P(g_i^N = 1|g_i^O = 1) = 1$, from which it follows that $P(g_i^N = 1, g_i^O = 1) = P(g_i^N = 1|g_i^O = 1)P(g_i^O = 1) = f_1^O$. To model this extreme scenario, that of no correlation ($P(g_i^N = 1|g_i^O = 1) = f^N$), and others in between, I assume that $P(g_i^N = 1|g_i^O = 1)$ can be written as a linear function of a parameter c (for correlation) that reflects the correlation between gene expression in the old and new states. I restrict this parameter to range from zero (uncorrelated expression) to one (perfectly correlated expression), because negative correlations are generally not observed in genome-wide gene expression measurements (File S6, File S7, Figure S1B and S1C). The linear function that fulfills the necessary constraints $P(g_i^N = 1|g_i^O = 1) = f^N$ for uncorrelated expression states ($c = 0$), and $P(g_i^N = 1|g_i^O = 1) = 1$ for perfectly correlated expression states ($c = 1$) is $P(g_i^N = 1|g_i^O = 1) = f^N + c(1 - f^N)$. One can show (File S6) that c is identical to the Pearson product-moment correlation coefficient R between gene expression states except for a linear scaling factor. I use c instead of R in all calculations below, because it leads to simpler mathematical expressions that are easier to interpret intuitively.

With this notation, equation (8) takes on the form

$$f_{11}^N = f_1^O(f^N + c(1 - f^N)) \tag{12a}$$
$$= (f_{11}^O + f_{01}^O)(f^N + c(1 - f^N)) \tag{12b}$$

The fraction of genes that are incorrectly on in the new state then follows from the relationship $P(g_i^N = 0|g_i^O = 1) = 1 - P(g_i^N = 1|g_i^O = 1) = 1 - (f^N + c(1 - f^N))$ as

$$f_{01}^N = f_1^O(1 - f^N - c(1 - f^N)) \tag{13a}$$
$$= (f_{11}^O + f_{01}^O)(1 - c)(1 - f^N), \tag{13b}$$

The fraction f_{00}^N of genes that are correctly off can be obtained as follows. Elementary probability theory dictates that

$$P(g_i^N = 1) = P(g_i^N = 1|g_i^O = 0)P(g_i^O = 0) \tag{14a}$$
$$+ P(g_i^N = 1|g_i^O = 1)P(g_i^O = 1) \tag{14b}$$

which is equivalent to

$$f^N = [1 - P(g_i^N = 0|g_i^O = 0)](1 - f_1^O) \tag{15a}$$
$$+ [f^N + c(1 - f^N)]f_1^O \tag{15b}$$
$$\tag{15c}$$

Solving this equation for $P(g_i^N = 0|g_i^O = 0)$ yields

$$P(g_i^N = 0|g_i^O = 0) = \frac{(1 - f^N)[1 - (1 - c)f_1^O]}{1 - f_1^O} \tag{16a}$$

and together with the identity $f_1^O = f_{11}^O + f_{01}^O$ I obtain

$$f_{00}^N = P(g_i^N = 0|g_i^O = 0)P(g_i^O = 0) \tag{17a}$$
$$= \frac{(1 - f^N)[1 - (1 - c)f_1^O]}{1 - f_1^O}(1 - f_1^O) \tag{17b}$$
$$= (1 - f^N)[1 - (1 - c)f_1^O] \tag{17c}$$
$$= (1 - f^N)[1 - (1 - c)(f_{11}^O + f_{01}^O)] \tag{17d}$$

From this expression, I use $P(g_i^N = 1|g_i^O = 0) = 1 - P(g_i^N = 0|g_i^O = 0)$ to obtain the fraction of genes that are incorrectly off

$$f_{10}^N = f^N - f_1^O[f^N(1 - c) + c] \tag{18a}$$
$$= f^N - (f_{11}^O + f_{01}^O)[f^N(1 - c) + c]. \tag{18b}$$

I note that $f_{00}^N + f_{01}^N + f_{10}^N + f_{11}^N = 1$, as required. Next I will use the preceding expressions to calculate the *change* in the fraction of correctly expressed genes in the new state, when misregulation is present in the old state, as opposed to when it is absent ($f_{01}^O = 0, f_{10}^O = 0$). To this end, I define $\Delta f_{11}^m = f_{11}^N - f_{11}^N|_{m^-}$, where the right-most term means that f_{11}^N (given by equation (12b)) is evaluated in the absence of misregulation in the old state. Notice that to do so, it is not sufficient to set $f_{01} = f_{10} = 0$ without also changing f_{11} and f_{00}, because otherwise the sum of these fractions will no longer equal one. To avoid specific assumptions about how f_{11} and f_{00} change, the calculations below use only the identity $f_1^O = f^O$, which must hold in the absence of misregulation. Together with equation (12b) and the identity $f_{11}^O = f^O - f_{10}^O$ from equation (1a), they yield

$$\Delta f_{11}^m = (f_{11}^O + f_{01}^O)[f^N + c(1 - f^N)] \tag{19a}$$
$$- (f^O)(f^N + c(1 - f^N)) \tag{19b}$$
$$= (f_{01}^O + f_{11}^O - f^O)[f^N + c(1 - f^N)] \tag{19c}$$
$$= \Delta_m[f^N + c(1 - f^N)] \tag{19d}$$

Here, $\Delta_m = f_{01}^O - f_{10}^O$ is the *excess* of wrongly active genes under misregulation. It is positive if there are more wrongly active genes than wrongly inactive genes. If $\Delta_m > 0$, misregulation is advantageous and this advantage grows with an increasing fraction of genes that need to be expressed in the new state, and with an increasing expression correlation c between the old and the new state.

I note that the derivation of equation (19) assumes that misregulation does not affect the conditional probabilities $P(g_i^N = 1|g_i^O = 1)$. To justify this assumption, I first note that the 'boundary conditions' $P(g_i^N = 1|g_i^O = 1) = f^N$ for uncorrelated expression states ($c = 0$), and $P(g_i^N = 1|g_i^O = 1) = 1$ for perfectly correlated expression states ($c = 1$) do not depend on the presence of misregulation. They follow from fundamental probability theory. In between these extremes, $P(g_i^N = 1|g_i^O = 1)$ must depend monotonically on the correlation c. The linear function $P(g_i^N = 1|g_i^O = 1) = f^N + c(1 - f^N)$ embodies the simplest and hence most parsimonious such monotonic dependency. This leaves the question whether the correlation c between expression states may differ substantially in the presence and absence of misregulation. This will not be the case provided that (i) the difference between the fraction of wrongly active genes and wrongly inactive genes is modest, which holds for empirically sensible parameters (Figure 2C, 'empirical'), and (ii) there is no correlation between genes that must be active for optimal adaptation in the new state and genes that are misregulated in the old state. The latter condition is also sensible, because the identity of the genes that must be expressed for optimal adaptation in the new state is determined by the environment, whereas the identity of misregulated genes in my model is determined by the mutational dynamics of binding sites in a genome.

With analogous assumptions, the effect of misregulation on the fraction of correctly off genes, $\Delta f_{00}^m = f_{00}^N - f_{00}^N|_{m^-}$ computes as

$$\Delta f_{00}^m = (1 - f^N)[1 - (1 - c)(f_{11}^O + f_{01}^O)] \tag{20a}$$

$$- (1 - f^N)[1 - (1 - c)f^O] \tag{20b}$$

$$= (f^O - f_{11}^O - f_{01}^O)(1 - c)(1 - f^N) \tag{20c}$$

$$= -\Delta_m(1 - c)(1 - f^N) \tag{20d}$$

To compute the overall change in the fraction of correctly expressed genes, we can add equations (19d) and (20d) to obtain the simple expression

$$\Delta f_{11}^m + \Delta f_{00}^m = \Delta_m[(2f^N - 1)(1 - c) + c] \tag{21}$$

Analogously, we get for the fractions of misexpressed genes

$$\Delta f_{01}^m := f_{01}^N - f_{01}^N|_{m^-} = \Delta_m(1 - c)(1 - f^N), \tag{22}$$

and

$$\Delta f_{10}^m := f_{10}^N - f_{10}^N|_{m^-} = -\Delta_m[f^N(1-c)+c]. \tag{23}$$

The preceding equations are necessary to compute the effect that misregulation has on organismal fitness in the new environment. To this end, it is simplest to calculate the ratio r_w of fitness values with and without misregulation, which yields

$$\begin{aligned}
r_w &:= \frac{w^N}{w^N|_{m^-}} && \text{(24a)}\\
&= \frac{(1-s_{01})^{Gf_{01}^N}(1-s_{10})^{Gf_{10}^N}}{(1-s_{01})^{Gf_{01}^N|_{m^-}}(1-s_{10})^{Gf_{10}^N|_{m^-}}} && \text{(24b)}\\
&= (1-s_{01})^{G(f_{01}^N - f_{01}^N|_{m^-})}(1-s_{10})^{G(f_{10}^N - f_{10}^N|_{m^-})} && \text{(24c)}\\
&= (1-s_{01})^{G\Delta_m(1-c)(1-f^N)}(1-s_{10})^{-G\Delta_m[f^N(1-c)+c]} && \text{(24d)}\\
&\approx \left[1 - Gs_{01}\Delta_m(1-c)(1-f^N\right] && \text{(24e)}\\
&\times \left[1 + Gs_{10}\Delta_m[f^N(1-c)+c]\right] && \text{(24f)}\\
&\approx 1 + G\Delta_m\left[s_{10}[f^N(1-c)+c] - s_{01}(1-c)(1-f^N)\right] && \text{(24g)}
\end{aligned}$$

The two approximations (24e,24g) rely on the assumption that s_{01} and s_{10} are small, and that terms involving $s_{01}s_{10}$ can be neglected. The fitness ratio takes an especially simple form when $s_{01} = s_{10} = s$, i.e.,

$$r_w \approx 1 + G\Delta_m s[(2f^N - 1)(1-c) + c] \tag{25}$$

If the ratio r_w exceeds one, then misregulation provides a net fitness advantage, i.e., misregulation in the old state increases fitness in the new state. If $s_{10} = s_{01} = s$, i.e., if selection acts equally strongly against incorrectly off genes than against incorrectly on genes, this will be the case when $\Delta_m > 0$ and when more than half of all genes are expressed in the new state, i.e., under the same conditions that increase the fraction of correctly expressed genes in (21).

References

Cardoso-Moreira, M., J. Halbert, D. Valloton, B. Velten, C. Chen, Y. Shao, A. Liechti, K. Ascencao, C. Rummel, S. Ovchinnikova, P. V. Mazin, I. Xenarios, K. Harshman, M. Mort, D. N. Cooper, C. Sandi, M. J. Soares, P. G.

Ferreira, S. Afonso, M. Carneiro, J. M. A. Turner, J. L. VandeBerg, A. Fallahshahroudi, P. Jensen, R. Behr, S. Lisgo, S. Lindsay, P. Khaitovich, W. Huber, J. Baker, S. Anders, Y. E. Zhang, and H. Kaessmann (2019). Gene expression across mammalian organ development. *Nature 571*(7766), 505–509.

Colbourne, J. K., M. E. Pfrender, D. Gilbert, W. K. Thomas, A. Tucker, T. H. Oakley, S. Tokishita, A. Aerts, G. J. Arnold, M. K. Basu, D. J. Bauer, C. E. Caceres, L. Carmel, C. Casola, J. H. Choi, J. C. Detter, Q. F. Dong, S. Dusheyko, B. D. Eads, T. Frohlich, K. A. Geiler-Samerotte, D. Gerlach, P. Hatcher, S. Jogdeo, J. Krijgsveld, E. V. Kriventseva, D. Kultz, C. Laforsch, E. Lindquist, J. Lopez, J. R. Manak, J. Muller, J. Pangilinan, R. P. Patwardhan, S. Pitluck, E. J. Pritham, A. Rechtsteiner, M. Rho, I. B. Rogozin, O. Sakarya, A. Salamov, S. Schaack, H. Shapiro, Y. Shiga, C. Skalitzky, Z. Smith, A. Souvorov, W. Sung, Z. J. Tang, D. Tsuchiya, H. Tu, H. Vos, M. Wang, Y. I. Wolf, H. Yamagata, T. Yamada, Y. Z. Ye, J. R. Shaw, J. Andrews, T. J. Crease, H. X. Tang, S. M. Lucas, H. M. Robertson, P. Bork, E. V. Koonin, E. M. Zdobnov, I. V. Grigoriev, M. Lynch, and J. L. Boore (2011). The ecoresponsive genome of daphnia pulex. *Science 331*(6017), 555–561.

Dragosits, M., V. Mozhayskiy, S. Quinones-Soto, J. Park, and I. Tagkopoulos (2013). Evolutionary potential, cross-stress behavior and the genetic basis of acquired stress resistance in Escherichia coli. *Molecular Systems Biology 9*, 643.

Eisenberg, E. and E. Y. Levanon (2013). Human housekeeping genes, revisited. *Trends in Genetics 29*(10), 569–574.

Gasch, A., P. Spellman, C. Kao, O. Carmel-Harel, M. Eisen, G. Storz, D. Botstein, and P. Brown (2000). Genomic expression programs in the response of yeast cells to environmental change. *Molecular Biology of the Cell 11*, 4241–4257.

Henry, G. L., F. P. Davis, S. Picard, and S. R. Eddy (2012). Cell type-specific genomics of Drosophila neurons. *Nucleic Acids Research 40*(19), 9691–9704.

Huang, D., W. Wu, S. R. Abrams, and A. J. Cutler (2008). The relationship of drought-related gene expression in arabidopsis thaliana to hormonal and environmental factors. *Journal of Experimental Botany 59*(11), 2991–3007.

Landis, G. N., D. Abdueva, D. Skvortsov, J. D. Yang, B. E. Rabin, J. Carrick, S. Tavare, and J. Tower (2004). Similar gene expression patterns characterize aging and oxidative stress in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America 101*(20), 7663–7668.

Uhlen, M., L. Fagerberg, B. M. Hallstroem, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjoestedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigyarto, J. Odeberg,

D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, and F. Ponten (2015). Tissue-based map of the human proteome. *Science 347*(6220), 1260419.

Wang, D., B. Eraslan, T. Wieland, B. Hallstrom, T. Hopf, D. P. Zolg, J. Zecha, A. Asplund, L.-h. Li, C. Meng, M. Frejno, T. Schmidt, K. Schnatbaum, M. Wilhelm, F. Ponten, M. Uhlen, J. Gagneur, H. Hahne, and B. Kuster (2019). A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Molecular Systems Biology 15*(2), e8503.